
Identification of context-dependent motifs by contrasting ChIP binding data.

Journal: Bioinformatics

Publication Year: 2010

Authors: Mike J Mason, Kathrin Plath, Qing Zhou

PubMed link: 20870645

Funding Grants: In vitro reprogramming of mouse and human somatic cells to an embryonic state

Public Summary:

DNA binding proteins play crucial roles in the regulation of gene expression. Transcription factors (TFs) activate or repress genes directly while other proteins influence chromatin structure for transcription. Binding sites of a TF exhibit a similar sequence pattern called a motif. However, a one-to-one map does not exist between each TF and motif. Many TFs in a protein family may recognize the same motif with subtle nucleotide differences leading to different binding affinities. Additionally, a particular TF may bind different motifs under certain conditions, for example in the presence of different co-regulators. The availability of genome-wide binding data of multiple collaborative TFs makes it possible to detect such context-dependent motifs. We developed a new motif finder, called CMF, for the de novo identification of motifs that are differentially enriched in two sets of sequences. Applying this method to a number of TF binding datasets from mouse embryonic stem cells, we demonstrate that CMF achieves substantially higher accuracy than several well-known motif finding methods. The software CMF is freely available for academic use at www.stat.ucla.edu/~zhou/CMF.

Scientific Abstract:

MOTIVATION: DNA binding proteins play crucial roles in the regulation of gene expression. Transcription factors (TFs) activate or repress genes directly while other proteins influence chromatin structure for transcription. Binding sites of a TF exhibit a similar sequence pattern called a motif. However, a one-to-one map does not exist between each TF and motif. Many TFs in a protein family may recognize the same motif with subtle nucleotide differences leading to different binding affinities. Additionally, a particular TF may bind different motifs under certain conditions, for example in the presence of different co-regulators. The availability of genome-wide binding data of multiple collaborative TFs makes it possible to detect such context-dependent motifs. **RESULTS:** We developed a contrast motif finder (CMF) for the de novo identification of motifs that are differentially enriched in two sets of sequences. Applying this method to a number of TF binding datasets from mouse embryonic stem cells, we demonstrate that CMF achieves substantially higher accuracy than several well-known motif finding methods. By contrasting sequences bound by distinct sets of TFs, CMF identified two different motifs that may be recognized by Oct4 dependent on the presence of another co-regulator and detected subtle motif signals that may be associated with potential competitive binding between Sox2 and Tcf3. **AVAILABILITY:** The software CMF is freely available for academic use at www.stat.ucla.edu/~zhou/CMF.

Source URL: <http://www.cirm.ca.gov/about-cirm/publications/identification-context-dependent-motifs-contrasting-chip-binding-data>